REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
31-05-2016

2. REPORT TYPE
FINAL REPORT

3. DATES COVERED *(From - To)*
12-02-2015 -- 31-05-2016

4. TITLE AND SUBTITLE
Topological Analysis of Wireless Networks (TAWN)

5a. CONTRACT NUMBER
HR0011-15-C-0050

5b. GRANT NUMBER
N/A

5c. PROGRAM ELEMENT NUMBER
N/A

6. AUTHOR(S)
Robinson, Michael

5d. PROJECT NUMBER
N/A

5e. TASK NUMBER
N/A

5f. WORK UNIT NUMBER
N/A

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
American University
Department of Mathematics and Statistics
4400 Massachusetts Ave NW
Washington, DC 20016

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Defense Advanced Research Projects Agency
675 N Randolph Road
Arlington, VA 22203-2114

10. SPONSOR/MONITOR'S ACRONYM(S)
DARPA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

13. SUPPLEMENTARY NOTES
N/A

14. ABSTRACT

The goal of this project was to develop topological methods to detect and localize vulnerabilities of wireless communication networks to jamming and traffic overload. Our analysis used high-dimensional cell complexes to describe broadcast resources. Local protocol, activity, and channel conditions can be associated to such a cell complex using a mathematical object called a sheaf. We leveraged the existing mathematical literature on sheaves that describes how to draw global (network-wide) inferences from them.

15. SUBJECT TERMS
Wireless network, local homology, sheaf, topology

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Michael Robinson
U	U	U	UU	32	19b. TELEPHONE NUMBER *(Include area code)* 202-885-3681

Final Report: May 2016

Topological Analysis of Wireless Networks

Principal Investigator:
Prof. Michael Robinson
Assistant Professor
Mathematics and Statistics
American University
Office: (202)885-3681
Mobile: (484)477-3345
michaelr@american.edu

Sponsor:
Dr. Wayne Phoel
Defense Advanced Research
Projects Agency
Strategic Technologies Office
(STO)
675 N. Randolph Street
Arlington, VA 22203-2114
(703)526-2724
wayne.phoel@darpa.mil

Program: TAWN Program
PR No. HR0011513049
Issued by DARPA/CMO under Contract No. HR0011-15-C-0050

The views, opinions and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

American University

The goal of this project was to develop topological methods to detect and localize vulnerabilities of wireless communication networks to jamming and traffic overload. Our analysis used high-dimensional *cell complexes* to describe broadcast resources. Local protocol, activity, and channel conditions can be associated to such a cell complex using a mathematical object called a *sheaf*. We leveraged the existing mathematical literature on sheaves that describes how to draw global (network-wide) inferences from them. We demonstrated that sheaf-based inferences can ascertain whether a wireless network is vulnerable to traffic overload and intentional jamming. Contrary to expectations, we found a significant *purely topological influence* on network performance, and spent most of our effort examining these closely. This program was made successful because it employed the right theoretical toolset, which we were able to implement in software. We also made substantial use of the pre-existing network simulation tool `ns2`.

On this program, we made several important accomplishments:

(1) We discovered protocol-independent *topological effects* on wireless network performance
(2) We discovered two *sheaf encodings* of traffic handling protocols
  (a) The *network activation sheaf*, which describes coarse network behavior
  (b) The *data payload sheaf*, which encodes routing protocol information
(3) We showed that the *local homology dimension* detects vulnerable nodes
(4) We discovered the *forwarded packet distribution*, a network summary correlated with the stalk dimension of local homology that could be derived from network hardware
(5) We showed that local homology correlates with forwarded packets, which explains why nodes with high local homology are at greater risk
(6) We implemented the first ever *relative simplicial homology* library
(7) We formulated a new conjecture about wireless network tomography using geometry and topology

We discern three classes of immediate next steps that leverage our accomplishments. The first is to continue developing sheaf encodings of network protocols and implementing them in our software sheaf library. Although we found that there is a protocol-independent, topological component to network topology, it has been well-established that network protocols have an important influence on network performance and vulnerability (for instance [35, 16]). On this program we developed sheaf encodings of network routing protocols, but did not devote much time to their analysis. It is natural to use these sheaf encodings of network routing protocols to assess the relative importance between network topology and protocol effects. We also recognize that there is a close connection between netflow data (which is routinely measured by network hardware) and projections of sections of activation or data payload sheaves, but did not study this closely enough to explicate it completely.

The second area of future work involves applying our techniques to higher fidelity network simulations. For instance, we did not simulate any scenarios in which the nodes were in motion, though all three sheaf models (local homology, network activation, and the data payload sheaves) support this kind of scenario. We also discovered that a wireless network's global topology impacts its response to bursts of traffic, but did not construct a statistically large enough set of data to completely characterize the effect.

Finally, we made a number of mathematical and numerical conjectures about the models we developed. In particular, we are most interested in studying the connection between network topology and network geometry. A limitation of the `ns2` simulation is that it does not model packet degradation due to signal loss. Although we exploited this efffect on this program – because this means that `ns2` simulations are *purely* topological – it would be more realistic to include signal degradation as well. This would require enriching our sheaf models to include geometric effects. There is no apparent obstacle to performing this encoding.

Several unmet challenges remain:

(1) All of our analyses were performed in an offline manner – after the simulations finished running. In most cases, the analyses did not run quickly, because homological calculations exhibit poor scaling properties. This has been extensively addressed in the computational topology literature for persistent homology and simplicial homology, but not for relative homology, which is needed in the computation of local homology. There are substantial theoretical hurdles in optimizing local homology computations, especially regarding distributed computation of local homology.

(2) Determining the link complex from observed network traffic. This is known to be hard even for wired networks, but wireless propagation physics may provide additional, relevant information.

(3) Now that this program has provided protocol-independent, and indeed traffic-independent, measures of local vulnerability, it opens the door for testing specific network attack and defense patterns. Aside from a few limited examples – simple distributed denial of service attacks – we did not perform a systematic analysis of network defense or attack strategies. For instnace, we did not address adaptively changing the network topology based on local homology, which seems a natural strategy for defense.

1. Task objectives

This project had the following objectives:

(1) Develop network health monitoring algorithms for use in simulation and laboratory experiments

(2) Detect and localize vulnerabilities of wireless communication networks to jamming and traffic congestion

(3) Assess the effectiveness of both offensive and defensive strategies for managing wireless communications in contested or busy environments

Our effort was divided across four technical tasks, namely:

Task 1: *Developing theory and algorithms*, in which we

(1) *Extended* the transmission sheaf model on the cell complexes described above to capture dynamic network behaviors including realistic protocols.

(2) *Explained* the validity of these models through illustrative examples.

(3) *Identified* topological invariants associated to the models that detect and localize network vulnerabilities.

(4) *Encoded* these invariants in practical algorithms that produce actionable information about network health.

Task 2: *Developing simulation* of wireless networks, in which we

(1) *Implemented* a simple, but realistic model of wireless network dynamics selected from those available in the literature.

(2) *Constructed* datasets using this network model under various traffic loads and adversarial jamming conditions.

Task 3: *Analyzing simulated data* using the algorithms and theory developed in Task 1, in which we

(1) *Applied* the algorithms to the data simulated in Task 2.

(2) *Characterized* their performance when used to detect network vulnerabilities.

(3) *Proved* theoretical guarantees about the detectibility of vulnerabilities using topological invariants.

Task 4: *Reporting*, in which we wrote technical reports summarizing our findings on Tasks 1-3.

1.1. **Task 1 milestones.**

1.1.1. *Static vulnerability assessment algorithm that inputs the state of the network at a single time.* We succesfully implemented the following algorithms:

(1) Assessment of global vulnerability using persistent homology

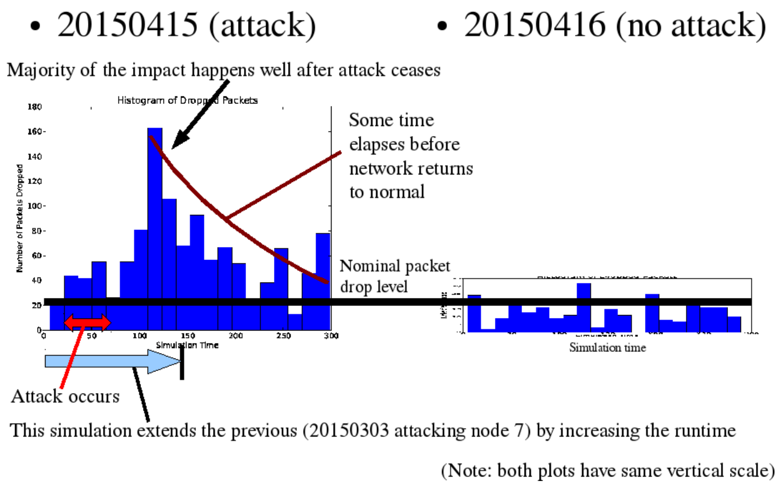
(2) Assessment of local vulnerability using local homology

Both of these algorithms ingest the network simulation models and simulated data as produced in support of Task 2.

1.1.2. *Sheaf model of a network that represents a simple store and forward protocol.* We successfully developed a sheaf encoding of a store and forward protocol and described this encoding in a report (See Section 5.1). When we initially proposed this effort, we expected that this kind of detailed protocol model would be necessary to obtain robust inferences about network vulnerability. Since the existing network vulnerability literature focuses most heavily on protocol-level behaviors, we expected that protocol effects were dominant. Contrary to expectations, we found a significant *purely topological influence* on network performance. Therefore, we devoted most of the rest of our effort on this program to studying the effects of local homology rather than more detailed protocols.

1.1.3. *An algorithm that detects vulnerabilities that evolve over time in the store-and-forward model.* An attack may have its greatest effect on the network well after it subsides. We ran a pair of simulations (See Figure 1; one with an attack and one without an attack) to analyze the transient behavior of the network post-attack. We found that the network does eventually return to normal after the attack subsides, and suspect that the time constant may be influenced by network topology.

1.1.4. *A sheaf model that represents a realistic media acess protocol.* When we initially proposed this milestone, we did not realize that a simpler topological invariant – namely the local homology invariant we found earlier in the project – would be as powerful as it appears to be. We have therefore stood down on developing algorithms based on the store-and-forward sheaf model. Indeed, we found that



- 20150415 (attack)
- 20150416 (no attack)

Majority of the impact happens well after attack ceases

Some time elapses before network returns to normal

Nominal packet drop level

Attack occurs

This simulation extends the previous (20150303 attacking node 7) by increasing the runtime

(Note: both plots have same vertical scale)

FIGURE 1. A network returns to normal after a burst of traffic

time-dependent vulnerabilities (as expressed in this milestone) can be detected using the local homology invariant alone. Since local homology is blind to protocol, we believe that dynamic vulnerabilities may be strongly governed by network topology even when protocol-specific effects are present.

We also found that in addition to the sheaf models we discovered, there is a cosheaf dual to the network activation sheaf. This cosheaf is rather natural, also models network usage, but does not exhibit the idle states of links.

1.1.5. *An algorithm that detects vulnerabilities using the realistic protocol model.* We successfully implemented our local homology invariant and deployed it against realistic network traffic.

1.1.6. *Construction of a sheaf-based network model and associated network monitoring algorithms.* As noted before, we found that local homology was more effective at describing network performance than we had expected. We therefore implemented an algorithm the measure the local topology of a network as a proxy for local node vulnerability.

1.2. **Task 2 milestones.**

1.2.1. *Implement a simple, but realistic model of wireless network dynamics selected from those available in the literature.* We had originally intended to develop paired simulations with and without attack patterns for different topologies. However, we found that network performance could be disrupted purely by increasing the number of links, without specific attacks. This turned out to be sufficiently rich that we have avoided modeling attacks in our more recent simulations. We have focused on three particular network topologies:

(1) A circular network – in which connections are confined to a thin boundary of disk
(2) A rectangular grid
(3) A small tree-like network

1.2.2. *Construct datasets using this network model under various traffic loads and adversarial jamming conditions.* Simulation data was successfully generated for this project, under the direction of three undergraduate student researchers: Danesh Krishnarao, Eyerusalem Abebe, and James Palladino. The following scenarios were constructed and are included in the final deliverable package:

Dataset name	Spatial layout of nodes	Attack pattern
20150129	Uniform random on square	None
20150211	Uniform random on square	Colocated nodes 50-53 attack
20150212	Uniform random on square	Node 50 attacks
20150224	Small fixed tree	Central and peripheral nodes attack
20150303	Small fixed tree	Peripheral nodes attack
20150415	Small fixed tree Longer runtime	Peripheral nodes attack
20150416	Small fixed tree Longer runtime	None
20150627	Rectangular networks Varying traffic	None
20150721	Rectangular networks Varying traffic	None
20150723	Rectangular networks Varying traffic	None
20150728	Circular networks Varying size and traffic	None

1.3. **Task 3 milestones.**

1.3.1. *Apply the algorithms developed in Task 1 to the data simulated in Task 2.* We have successfully performed analyses of the simulation data. These analyses included a cursory traditional packet loss analysis, for use as “ground truth” regarding the state of the network in our topological analysis. We later found a novel *forwarded packet distribution* (Section 4.4) that incorporates both topological and traffic effects.

We demonstrated the usefulness of two topological invariants on the simulated network data:

(1) Persistent homology given the known location of the attackers
(2) Local homology of the link complex of the network.

We found that the persistent homology invariant was able to coarsely identify when the network was more vulnerable to a particular attack pattern. The local homology invariant appears to give much more refined vulnerability estimates and we spent most of our subsequent effort analyzing it.

To validate our results, we took several networks with different global topological structure (as measured by persistent homology, Section 4.1.2), and analyzed the time constant for returning to normal conditions after an attack subsides. We examined packet drop rates as a proxy for network health, focusing on the connection to nontrivial higher homology groups of the network’s link complex.

1.3.2. *Characterize the performance of the algorithms when used to detect network vulnerabilities.* We proposed the following hypothesis: *nontrivial loops in the network can provide robustness at the cost of some extra latency.* To test this hypothesis, we constructed a particular network (20150728) in which the nodes were laid along a circle. Since `ns2` does not simulate noise, there is no difference in simulation between two nearby nodes and two far apart nodes provided both sets of nodes are connected. By varying the radius of the circle on which the nodes lie, we can isolate the study of the network's topology from geometric effects. The link complex of the network is either topologically trivial (trivial H_1) or homotopic to a circle (nontrivial H_1). We randomized the communication links to be established by the network with different radii and analyzed the packet counts. We found that networks with nontrivial topology appear to result in more forwarded packets, but less packet drops. The overall traffic carried by the network ends up being *higher* with a more strongly connected network due to the need for resends. We performed a systematic study (20150211, 20150224, and 20150728) of packet loss as a function of topology. We considered the number of packets forwarded, dropped, and acknowledged by all nodes under independent variations of network connectivity and traffic levels. We are now certain that global network topology (in terms of nontrivial homology) plays a central role in determining overall performance.

It is known that local homology can be used for finding the boundary of manifolds, and it appears to change near the "boundary" of the link complex. We wondered if there is a reflection of this in the traffic of the network as well. Nodes along the periphery of the network tend to forward fewer packets. Therefore, it appears that the dimension of local homology groups and forwarded packet counts are correlated. We began studying this phenomenon, and discovered a good way to characterize it is using a probabilistic description called the *forwarded packet distribution.* Specifically, we found that in a dense network organized as a rectangular grid, relatively few nodes forward most of the traffic. These nodes appear to be away from the topological boundary of the network as determined by local homology. As might be expected, the distribution of forwarded packet counts is rather dependent on the particular traffic being handled by the network.

We spent a little time studying the local homology sheaf of the whole network versus time snaphots of the network: attempting to answer the question "How long do you need to observe the network to get its topology right?" It appears that without further constraints (such as the curvature constraint mentioned earlier) this problem cannot be solved directly. Only the strongest constraints – for instance, all links at all non-leaf nodes must be exercised – appear to yield precise performance guarantees [35].

1.3.3. *Prove theoretical guarantees about the detectibility of vulnerabilities using topological invariants.* We proved a precise, relative homological bound on the number of components a network is split into under the influence of interference. This provides a precise theoretical justification for why local homology is an important determinnant of network performance.

We spent considerable time looking at the correlation of local homology with packet counts. Although this does not specifically address dynamic behaviors, we formed the hypothesis that an exact sequence of sheaves modulates our ability to sense the network from a smaller number of nodes. Indeed, such *netflow* data is often available from hardware, if not very rich. We proved that netflow is essentially uninteresting from a topological standpoint, although it may be useful statistically. (Actually, it is already known that netflow is statistically useful.) Due to the fact that our simulations give us routing details on the packets being sent, it may be possible to compute how much information is available from netflow versus the richer data provided by our simulation.

The local homology invariants we studied are related to *geometric* structure – namely curvature – via the Gauss-Bonnet theorem. Since it's well-known that the Gauss-Bonnet theorem holds in geometric realizations of triangulated surfaces, the question is exactly how well it works in our network models. Based on some preliminary analyses, we believe that a version of the Gauss-Bonnet theorem holds in the case of link complexes. Tantalizingly, if the curvature of a surface is uniformly negative, then the Radon transform becomes injective. It is therefore possible to perform *lossless tomography* in smooth hyperbolic manifolds. We conjecture that a similar condition may exist on networks, and would therefore provide conditions under which lossless network tomography is possible. Intuitively, this is the case of expander graphs, in which the boundary of a set of nodes is much larger than the set's interior.

Finally, although the space of global sections for an activation sheaf is a useful invariant, and is directly helpful for specifying the neighborhoods in the local homology invariant, we proved that the vectorized formulation of the activation sheaf is acyclic. It does not contain any further algebraic invariants that could be exploited (Theorem 28). This came as somewhat of a surprise, since the same trick of vectorifying a sheaf of sets produced interesting invariants on a previous project [32].

1.4. **Task 4 milestones.** See Section 5.1 for a list of publications and presentations that resulted from this project.

2. Technical problem statement

The goal of the this program was to develop topological methods to *detect and localize vulnerabilities* of wireless communication networks to jamming and traffic overload. Current methods rely on graph-based models, which study the vulnerabilities of pairwise links between nodes. This is appropriate for wired networks, but cannot treat broadcast resources effectively. Additionally, network protocols are described in an *ad hoc* manner that impedes the analysis of large networks.

Instead of graphs, our analysis uses high-dimensional *cell complexes* (of which graphs are a special case) to describe broadcast resources. Local protocol, activity, and channel conditions can by associated to such a cell complex using a mathematical object called a *sheaf.* There already exists a substantial mathematical literature on sheaves that describes how to draw global (network-wide) inferences from them. Our initial analysis indicates that sheaf-based inferences can ascertain whether a wireless network is vulnerable to traffic overload and intentional jamming.

Failure of a critical wireless communication network can severely hamper operation of our military or commericial infrastructures. Very few current methods exist for assessing hidden vulnerabilities of communication networks. Indeed, the reliability of wireless communication networks is usually experimentally determined rather than theoretically proven. The techniques we developed will allow network operators to *uncover hidden vulnerabilities* and *assess the effectivness of countermeasures*. Additionally, our approach could provide a technique for assessing the effectiveness of both *offensive and defensive strategies* for managing wireless communications, which supports electronic warfare operations in contested areas.

2.1. **Historical Context.** Following [31, 21], we will declare that a network is *vulnerable* at a collection of links if their removal results in a disconnected network. Although this is a fairly drastic mode of failure (the loss of a few links can dramatically reduce the network capacity [3] without disconnecting the network), it is a failure that is easy to describe.

Although vulnerability to purely random failures has been extensively studied, Commander *et al.* [8] observe that vulnerability to an adversarial jammer has received little attention. Xu *et al.* [36] was one of the few works discussing jammers, from both the perspective of the attacker and defender.

Graph theory plays a central role in identifying *critical nodes* [3] – those through which a substantial amount of traffic passes. Because identifying these nodes is computationally difficult [12, 13], our (more naïve) definition of vulnerability as causing disconnectedness avoids a combinatorial explosion. Although the use of graph theory is well-established in studying resource conflict in wireless networks [30, 38, 23, 26], we are inspired by the detailed survey [6], which states, "...problems over networks with randomly varying topology remains an under-explored area with little known results on models or methodologies."

Recently, vulnerability assessments [17, 14, 34, 4] have been performed successfully on networks with hierarchical structure using percolation theory. The earliest such paper on the subject [17] found that network topology is a performance driver for network robustness, since the property of *connectedness* plays an important role in their activation model. Specifically, nodes need to be connected to their local source else they deactivate. This is a good model for utility distribution networks, but is not consistent with *ad hoc* wireless network usage of broadcast resources.

The tools of topology have been used more extensively in sensor networks. Our simplicial complexes are inspired by the work of [11, 20, 10, 25]. Jamming and defense problems for sensor networks were discussed in [2, 37, 39]. Finally, we observe that communication network capacity has been successfully studied [19] using the tools of sheaf theory.

3. General methodology

3.1. **Abstract simplicial complexes.** Take the number of path components of a network as a measure of its health: fewer components is better. If there are multiple path components, there exist pairs of nodes which cannot communicate with each other, even through relays. We also make the following *single channel assumption*: if a link connected to a node is jammed, then that node cannot receive transmissions from *any* other node.

Definition 1. A wireless network *vulnerability* is its susceptibility to becoming disconnected when a single source of interference is present.

Central to our approach to assessing vulnerabilities of this form is the use of cell complexes, such as the following:

Definition 2. An *abstract simplicial complex* X on a set A is a collection of ordered subsets of A that is closed under the operation of taking subsets. We call an element of X which itself contains $k+1$ elements a *k-cell*. We usually call a 0-cell a *vertex* and a 1-cell an *edge*.

If a, b are cells with $a \subset b$, we say that a is a *face* of b, and that b is a *coface* of a. A cell of X that has no cofaces is called a *facet*.

The *closure* $\text{cl}\ S$ of a set S of cells in X is the smallest abstract simplicial complex that contains S. The *star* $\text{star}\ S$ of a set S of cells in X is the set of all cells that have at least one face in S.

Suppose a radio network consisting of a collection of nodes $N = \{n_i\}$ is active in a spatial region R. Assume all nodes communicate through a single-channel, broadcast resource. An open set $U_i \subset R$ is associated to each node n_i that represents its *transmitter coverage region.* (See Figure 2.) For each node n_i, a continuous function $s_i : U_i \to \mathbb{R}$ represents its *signal level* at each point in U_i. Without loss of generality, we assume that there is a global threshold T for accurately decoding the transmission from any node.

Definition 3. The *interference complex* $I = I(N, U, s, T)$ consists of all subsets of N of the form $\{i_1, \dots, i_n\}$ for which $U_{i_1} \cap \cdots \cap U_{i_n}$ contains a point $x \in R$ for which $s_{i_k}(x) > T$ for all $k = 1, \cdots n$.

Briefly, the interference complex describes the lists of transmitters that when transmitting will result in at least one mobile receiver location receiving multiple signals simultaneously. (Without the constraint on the decoding threshold, the interference complex reduces to the well-known Čech complex [22].)

Proposition 4. *Each facet of the interference complex corresponds to a maximal collection of nodes that mutually interfere.*

Definition 5. The *link graph* is the following collection of subsets of N:

(1) $\{n_i\} \in N$ for each node n_i, and
(2) $\{n_i, n_j\} \in N$ if $s_i(n_j) > T$ and $s_j(n_i) > T$.

The *link complex* $L = L(N, U, s, T)$ is the clique complex of the link graph, which means that it contains all elements of the form $\{i_1, \dots, i_n\}$ whenever this set is a clique in the link graph.

Proposition 6. *Each facet in the link complex is a maximal set of nodes that can communicate directly with one another (with only one transmitting at a time).*

Corollary 7. *Facets of the interfence and link complexes represent common broadcast resources.*

Example 8. Figure 2 shows a network with three nodes. The coverage regions are shown at left for a given threshold T. Since there are nonempty pairwise intersections between the coverage regions, but there is no common point of intersection for all three nodes, the interference complex (middle) contains no 2-cells. However, no pair of nodes can actually communicate, so the link complex consists of three isolated vertices.

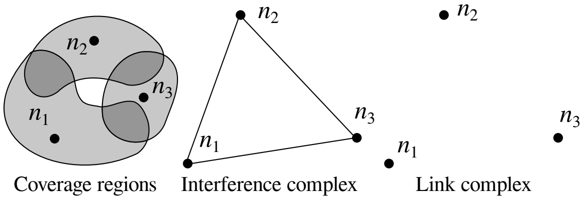


FIGURE 2. Transmitter coverage regions (left), the associated interference complex (middle), and link complex (right)

3.2. **Relative and local homology.** Suppose that $Y \subseteq X$ is a subcomplex of an abstract simplicial complex. The *relative k-chain space* $C_k(X,Y)$ is the abstract vector space[1] whose basis consists of the k-dimensional faces of X that are not in Y. We also write $C_k(X)$ in place of $C_k(X,\emptyset)$. Given these spaces, we can define the *relative boundary map* $\partial_k : C_k(X,Y) \to C_{k-1}(X,Y)$ given by

$$\partial_k(v_0,\dots,v_k) = \sum_{i=0}^{k}(-1)^i \begin{cases}(v_0,\dots,v_{i-1},v_{i+1},\dots,v_k) & \text{if } (v_0,\dots,v_{i-1},v_{i+1},\dots,v_k) \notin Y,\\ 0 & \text{otherwise}\end{cases}$$

Proposition 9. *(Completely standard, for instance see* [22, Lemma 2.1]*) The sequence of linear maps $(C_\bullet(X,Y),\partial_\bullet)$ is a chain complex.*

Definition 10. If $Y \subseteq X$ is a subcomplex of an abstract simplicial complex, then $H_k(X,Y) = H_k(C_\bullet(X,Y),\partial_\bullet)$ is called the *relative homology of the pair* (X,Y). We usually write $H_k(X) = H_k(X,\emptyset)$, which is the *simplicial homology* of X.

Proposition 11. [22, Prop. 2.10] *$H_k(X)$ is homotopy invariant: a homotopy equivalence $X \to Y$ between two abstract simplicial complexes induces isomorphisms on H_k.*

Proposition 12. [22, Props. 2.9, 2.19] *Each continuous function $f : X \to Z$ from one abstract simplicial complex to another which restricts to a continuous $Y \to W$ induces a linear map $H_k(X,Y) \to H_k(Z,W)$ for each k.*

Definition 13. (compare [22, end of Sec. 2.1], [27]) For an open subset $U \subseteq X$ of an abstract simplicial complex, the *local homology* at U is $H_k(X, X\backslash U)$.

Proposition 14. *(Excision for abstract simplicial complexes, compare* [29]*) If U is an open set of an abstract simplicial complex X, then $H_k(X, X\backslash U) \cong H_k(cl\ U, fr\ U)$ where $fr\ U = cl\ U \cap cl\ (X\backslash U)$ is the* frontier *of the set U.*

As an aside, we note that this is somewhat stronger than the usual exision principle, a usual formulation of which reads:

Proposition 15. *(Excision principle,* [22, Thm. 2.20]*) If U and V are sets in a topological space X for which $cl\ V \subseteq X\backslash(cl\ U)$, then $H_k(X, X\backslash U) \cong H_k(X\backslash V, X\backslash(U \cup V))$.*

We obtain Proposition 14 by taking U as an open set as before and $V = \text{int } X\backslash U$. Notice that this choice of V violates the hypotheses of Proposition 15 because $\text{cl } V = X\backslash U$ which is not generally a subset of $X\backslash(\text{cl } U)$.

[1] Since the software presented in later sections works over $\mathbb{R}$ vector spaces, we avoid the obvious generalization to modules over some ring.

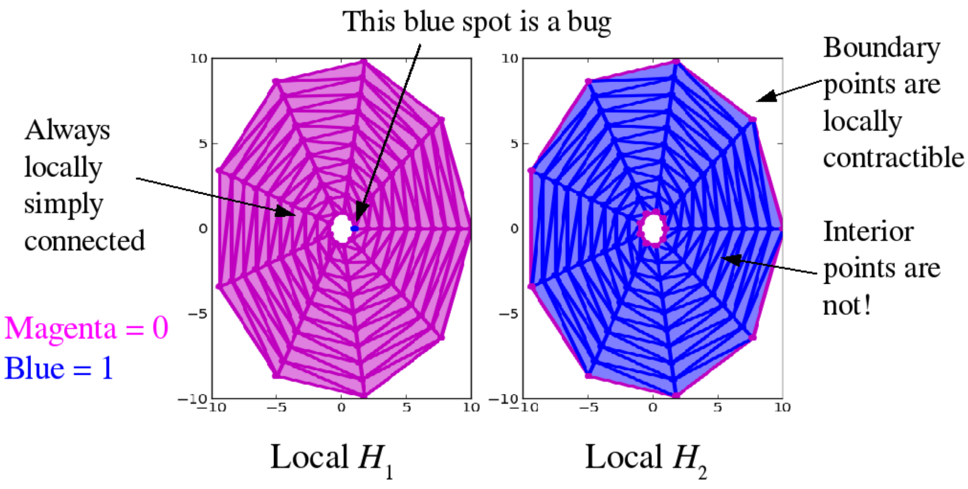


FIGURE 3. The local homology dimension for an annulus. Notice that the local homology of the interior is distinct from the boundary

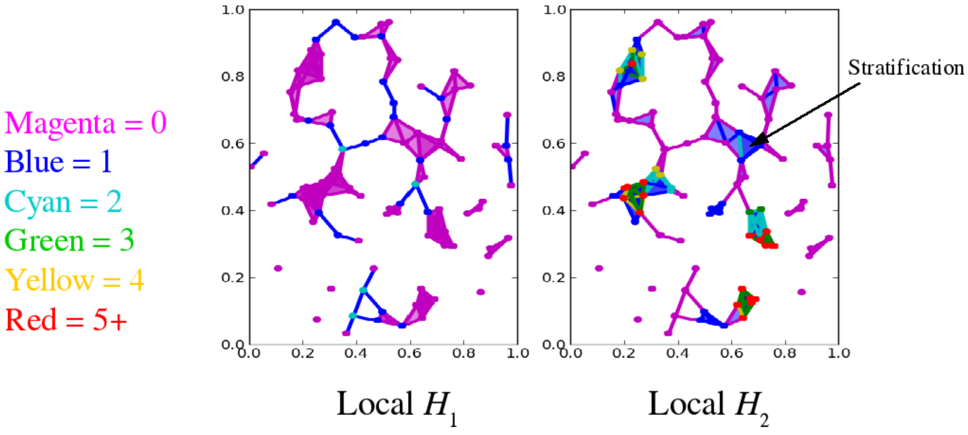


FIGURE 4

In this report, we shall assume that all abstract simplicial complexes are *locally finite*, which means that there is a finite open set containing each face. Then Proposition 14 indicates that local homology can be computed using finite dimensional linear algebra. Most of the computations we present in later sections use $U = \text{star } A$, which further limits the size of the computation.

Proposition 16. *The functor $U \mapsto H_k(X, X\backslash U)$ defines a sheaf; called the* sheaf of local k-homology.

Corollary 17. *Global sections of the local homology sheaf are the reduced homology classes of the abstract simplicial complex.*

Because of the excision principle, the sheaf of local k homology is a purely local invariant of a topological space. As Figure 3 shows, it is particularly useful for detecting the boundary of topological spaces. As Figure 4 shows, it generalizes the notion of the *degree of a vertex* in a graph to all faces of a simplicial complex. Our wider team (not on this project) has shown that local homology is therefore a useful general network science invariant [24].

Returning to the case of a wireless network, the single channel assumption means that an attack on a facet L removes its region of influence from the network. When this occurs, the nodes that are faces of this facet cannot communicate, but other portions of the network may become disconnected, too. A preliminary result is the following Theorem, which strongly suggests the value of using homological invariants to detect vulnerability.

Theorem 18. *Suppose that X is either a link or interference complex, and that L is a facet of X. If X is connected and roi L = star cl L is a proper subset of X, the number* rank $H_1(X, X \setminus \textit{roi } L) + 1$ *is an upper bound on the number of connected components that an attack on L cuts the network into. When $H_1(X)$ is trivial, that upper bound is attained.*

3.3. **Network simulations.** We constructed our network simulations using the `ns2` [1] simulation tool. We configured several wireless networks using the 802.11b protocol. Each node in the network was given the same power level, and was assigned a static location for the entire simulation. Empirically, we found that the typical connection radius was around 15 meters, which we used to control the topology of the networks under study. Despite the fact that node locations were specified, the simulations are purely topological in nature since `ns2` does not simulate bit error rates or other signal degradations. Given a network setup, traffic was then overlaid by specifying source and destination nodes as well as a payload. To keep the analysis simple, all transmissions consisted of `ftp` connections with varying data payloads.

The `ns2` simulator is controlled by the TCL scripting language. Therefore, simulations were specified as TCL scripts containing node and traffic setup. Since we often wanted to systematically vary some parameters of the simulation, we often wrote a python wrapper script that repeatedly called the `ns2` TCL script with command line arguments. Once completed, the `ns2` simulation produces a *transcript* file, which is a plain text file in which each line describes a particular packet transmission event. The format of the transcript file is easily parsed for analysis. In our analysis, we found it sufficient to filter the transcript files using the standard Unix `grep` utility with a small regular expression.

4. Technical results

The following sections detail our key findings and accomplishments as listed in the introduction.

4.1. **Protocol-independent global topological effects on network performance.** Our first key finding is that although traffic conditions and protocol do impact the vulnerability of a network (we are not refuting [35]!), there are significant global topological effects. We found that the presence of nontrivial topological

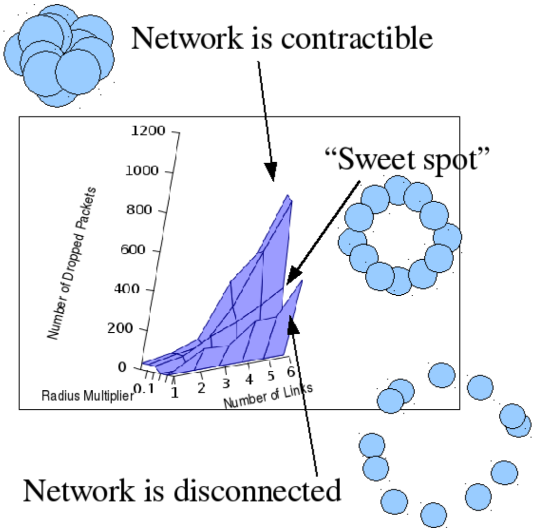


FIGURE 5. Dropped packets in a circular network (20150728) with varying radius and number of active links

features (loops, voids, etc.) in the link complex are generally protective. This appears to be because there are ample alternate routing paths in the network that remain even when the network is under attack.

4.1.1. *Preliminary circular network.* The simplest nontrivial topological space is a circle, which although connected is not simply connnected. Dataset 20150728 consists of a family of circular networks with a fixed number of nodes placed along circles of various radii and with various levels of traffic. Since the connection radius around each node is fixed, the link complex of the network goes through several topological transitions:

(1) When the radius of the network is large, none of the nodes can communicate, so the link complex consists of a set of vertices with no links,
(2) When the radius of the network is small, all of the nodes can communicate directly to one another, so the link complex consists of a single high-dimensional simplex (and all of its faces),
(3) Finally, when the radius of the network is between these limits, the link complex contains a nontrivial loop.

We simulated a varying number of connections, with a maximum of six connections. To help randomize over traffic conditions, we ran several simulations at each radius and number of links, but permuted the node identities while keeping the connections fixed.

We hypothesized (correctly) that settings (1) and (2) result in poor network performance, but for different reasons. In setting (1), clearly no communication is possible. In setting (2), although communication is possible, collisions dominate, since there is significant competition for the single shared broadcast channel. Therefore, it would appear that setting (3) is the best situation. Figure 5 summarizes our results, and shows the number of dropped packets of the network, as the network radius and number of links varies. As the number of links increases, the number of dropped packets increases and the network reaches saturation. A smaller radius means that more dropped packets as well, due to collisions on shared broadcast channels.

The presence of a topological loop appears to reduce network vulnerability to self-interference. There is a small increase in latency due to the need for forwarding in a topologically nontrivial network, however this is strongly outweighed by the latency penalty due to collisions in a topologically trivial network. This experiment indicates a strong performance difference driven by higher dimensional topology between wired networks and wireless networks.

4.1.2. *Persistent homology validation.* The previous experiment was rather simplistic and is based on the presence or absence of a loop in the network. Since loops can be measured by homology (Section 3.2), it seems like homology would be a good indicator of network robustness. However, homology is not robust to statistical variation. For this reason, persistent homology [15, 18, 7] was invented to improve its statistical robustness. We made use of the pre-existing `Perseus` software tool [28] to perform persistent homology calculations. Persistent homology analyzes a *filtration* of abstract simplicial complexes

$$X_1 \subseteq X_2 \subseteq \cdots \subseteq X_N$$

and computes a sequence of induced maps on the homology spaces

$$H_k(X_1) \to H_k(X_2) \to \cdots \to H_k(X_N)$$

identifying the nontrivial vectors in this sequence that persist across many spaces. As described in our preliminary work [33], persistent homology has value for detecting network vulnerability. On this project, we extended this result according to the process shown in Figure 6. Since the most natural way to use persistent homology for predicting the significance of a network disruption involves forming a "reverse" filtration in which it is more natural to reverse the indices $(1, \ldots, N)$. Therefore, as shown in Figure 6, we correct this reversal before passing the link complexes to `Perseus`.

Briefly, the persistent homology invariant for a wireless network stipulates network node locations and a progressively degrading network, usually due to the presence of some "attacker" nodes. We simply used the distance of a node to the nearest attacker node as a measure for how disrupted that particular node would be. Therefore, the impact of different attack patterns can be assessed. Figures 7 – 9 show several examples of networks with both the persistent homology invariant computed and a time-dependent packet loss histogram.

We computed two separate persistence diagrams, one for H_0 (connected components) and one for H_1 (loops). Each diagram is a multiset[2] of points in the plane. Each point in the diagram corresponds to a topological feature, either a connected component or a loop in the Figures. The importance of a topological feature is its distance from the main diagonal – points farther from the diagonal are more important. One should note that `Perseus` returns -1 for features that are maximally important, so these appear below the diagonal.

4.2. **Sheaf encodings of traffice handling protocols.**

[2] A *multiset* is a set in which duplications are permissible.

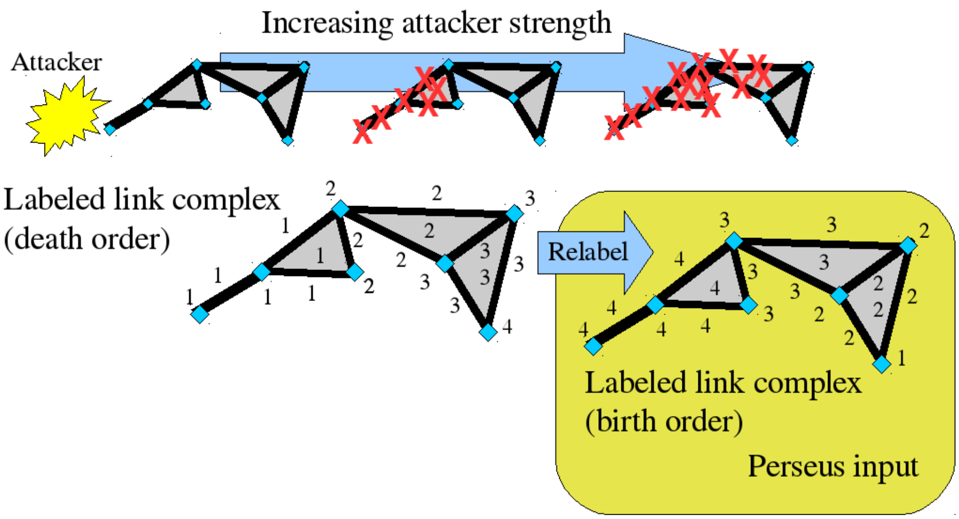


FIGURE 6. Process of transformating node and attacker locations into input for the Perseus persistent homology utility

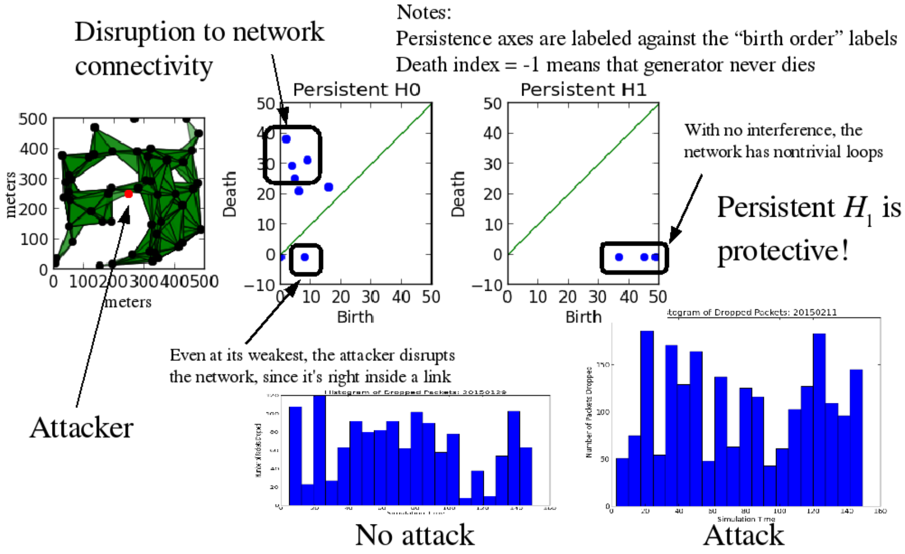


FIGURE 7. Results of the 20150211 network; no significant impact due to this attack pattern

4.2.1. *Network activation sheaves.* The interference caused by a transmission impacts the usability of the network outside of the transmission's immediate vicinity. This section builds a consistent definition of the *region of influence* of a node or a link within the network. To justify this definition, we use a local model that describes which configurations of nodes can transmit simultaneously.

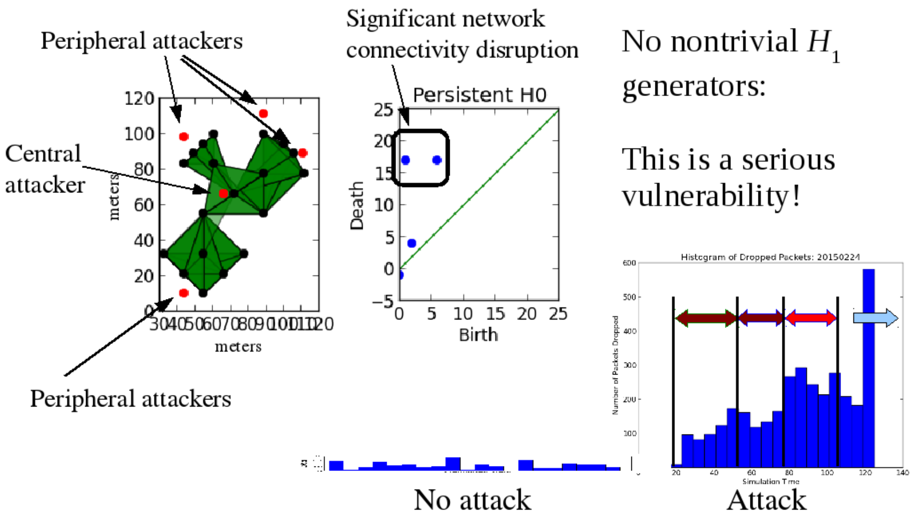


FIGURE 8. Results of the 20150224 network; significant network disruption due to attack

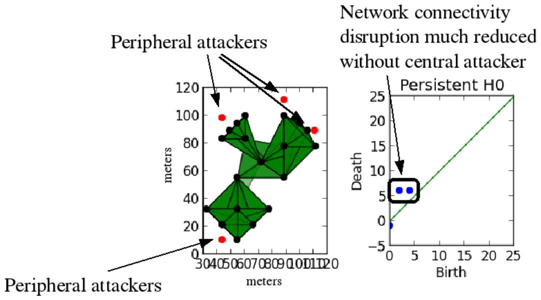


FIGURE 9. Results of the 20150303 network; no significant risk detected using persistent homology. (See also Figure 12)

Definition 19. Suppose that X is a simplicial complex (such as an interference or link complex) whose set of vertices is N. Consider the following assignment $\mathcal{A}$ of additional information to capture which nodes are transmitting and decodable:

(1) To each cell $c \in X$, assign the set

$$\begin{aligned}\mathcal{A}(c) \quad &= \quad \{n \in N : \text{there exists a cell } d \in X \text{ with} \\ &\qquad c \subset d \text{ and } n \in d\} \cup \{\perp\}\end{aligned}$$

of nodes that have a coface in common with c, along with the symbol $\perp$. We call $\mathcal{A}(c)$ the *stalk* of $\mathcal{A}$ at c.

(2) To each pair $c \subset d$ of cells, assign the *restriction function*

$$\mathcal{A}(c \subset d)(n) = \begin{cases} n & \text{if } n \in \mathcal{A}(d) \\ \perp & \text{otherwise} \end{cases}$$

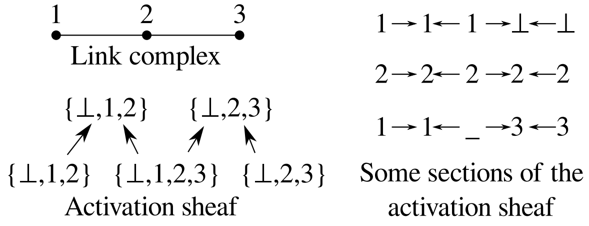


Figure 10. A link complex (left top), sheaf $\mathcal{A}$ (left bottom), and three sections (right). The restrictions are shown with arrows. Global section when node 1 transmits (right top), global section when node 2 transmits (right middle), and a local section with nodes 1 and 3 attempting to transmit, interfering at node 2 (right bottom)

For instance, if $c \in X$ is a cell of a link complex, $\mathcal{A}(c)$ specifies which nearby node is transmitting and decodable, or $\perp$ if none are. The restriction functions relate the decodable transmitting nodes at the nodes to which nodes are decodable along an attached wireless link. Similarly, if $c \in X$ is a cell of an interference complex, $\mathcal{A}(c)$ also specifies which nearby node is transmitting, and effectively locks out any interfering transmissions from other nodes.

Definition 20. The assignment $\mathcal{A}$ is called the *wireless activation sheaf* and is an example of a *cellular sheaf* – a mathematical object that stores local data. The theory of sheaves explains how to extract consistent information, which in the case of networks consists of nodes that whose transmissions do not interfere with one another.

A *section* of $\mathcal{A}$ supported on a subset $Y \subseteq X$ is an assignment $s : Y \to N$ so that for each $c \subset d$ in Y, $s(c) \in \mathcal{A}(c)$ and $\mathcal{A}(c \subset d)\,(s(c)) = s(d)$. A section supported on X is called a *global section*.

Specifically, global sections are complete lists of nodes that can be transmitting without interference.

Example 21. Figure 10 shows a network with three nodes, labeled 1, 2, and 3. When node 1 transmits, node 2 receives. Because node 2 is busy, its link to node 3 must remain inactive (right top). When node 2 transmits, both nodes 1 and 3 receive (right middle). The right bottom diagram shows a local section that cannot be extended to the cell marked with a blank. This corresponds to the situation where nodes 1 and 3 attempt to transmit but instead cause interference at node 2.

Example 22. Observe that in either of the simplicial complex models shown in Figure 2, only one of the nodes may transmit at a time.

Definition 23. Suppose that s is a global section of $\mathcal{A}$. The *active region* associated to a node $n \in X$ in s is the set

$$\text{active}(s, n) = \{a \in X : s(a) = n\},$$

which is the set of all nodes that are currently waiting on n to finish transmitting.

Corollary 24. *If s is a global section of an activation sheaf $\mathcal{A}$, then the* support *of s – the set of cells c where $s(c) \neq \perp$ – consists of a disjoint union of active regions of nodes.*

Lemma 25. *The active region of a node is independent of the global section. More precisely, if r and s are global sections of $\mathcal{A}$ and the active regions associated to $n \in X$ are nonempty in both, then $active(s, n) = active(r, n)$.*

Corollary 26. *The space of global sections of an activation sheaf consists of all sets of nodes that can be transmitting simultaneously without interference.*

Although the space of global sections for an activation sheaf is a useful invariant, its sheaf cohomology is rather uninteresting. We need to enrich their structure somewhat to see this, though.

Definition 27. If $\mathcal{A}$ is an activation sheaf on an abstract simplicial complex X, the *vector activation sheaf* $\widehat{\mathcal{A}}$ is given by specifying its stalks and restrictions:

(1) To each cell $c \in X$, let $\widehat{\mathcal{A}}(c)$ be the vector space whose basis is $\mathcal{A} \backslash \{\perp\}$ (so the dimension of this vector space is the cardinality of $\mathcal{A}$ without counting $\perp$)

(2) The restriction map $\widehat{\mathcal{A}}(c \subset d)(n)$ is the basis projection, which is well-defined since $\mathcal{A}(d) \subseteq \mathcal{A}(c)$.

Theorem 28. *The dimension of the cohomology spaces of a vector activation sheaf $\widehat{\mathcal{A}}$ on a link complex X are*

$$\textit{dim } H^k(\widehat{\mathcal{A}}) = \begin{cases} \textit{the total number of nodes} & \textit{if } k = 0 \\ 0 & \textit{otherwise} \end{cases}$$

4.2.2. *Data payload sheaves.* (Note that this is detailed more thoroughly in a separate technical report "Modeling wireless network routing using sheaves" that is included with the final deliverable.)

The activation sheaf describes the state of the network at a single instant in time. Because the network conditions may change over time, the link and interference complexes may also change with time. This section describes a general framework for representing both these changes and the data that is transmitted over the links.

In order to capture changes in the network's topology over time, it is appropriate to use a single link or interference complex to represent a single timeslice. To represent how the network's state evolves over several consecutive timeslices, we construct additional links between nodes in different timeslices. These links carry information from one timeslice to the next.

Extending the definition of a link complex above, again suppose that a radio network consists of a collection of nodes $N = \{n_i\}$ in a spatial region R in which a coverage region $U_i \subset R$ is associated to each node n_i. For each node n_i, assign a signal level function $s_i : U_i \times \mathbb{Z} \to \mathbb{R}$, where the second input represents time. Again, without loss of generality, we assume that there is a global decoding threshold T.

Definition 29. The *time-dependent link graph* is the following collection of subsets of $N \times \mathbb{Z}$:

(1) $\{(n_i, t)\} \in N$ for each node n_i and $t \in \mathbb{Z}$,

(2) $\{(n_i, t), (n_j, t)\} \in N$ if $s_i(n_j, t) > T$ and $s_j(n_i, t) > T$, and

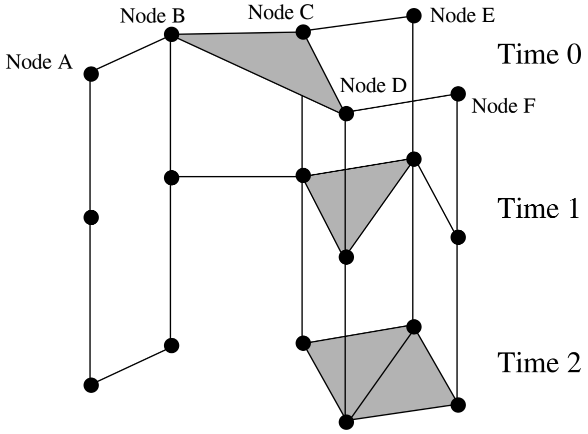


FIGURE 11. Evolution of a simplicial complex model of a wireless network through time. The nodes A–F listed at time 0 are repeated through the other timeslices as proceeding vertically downward through the diagram.

(3) $\{(n_i, t), (n_i, t+1)\}$ for each node n_i.

The *time-depdendent link complex* $L = L(N, U, s, T)$ is the clique complex of the time-dependent link graph, which means that it contains all elements of the form $\{i_1, \dots, i_n\}$ whenever this set is a clique in the link graph. The *time t timeslice* of L is the maximal subcomplex of L containing vertices from $N \times \{t\}$. The *time-dependent interference complex* and its timeslices can be defined in an analogous manner.

Figure 11 shows an example of a time-dependent link complex. Notice that each timeslice is a link complex, and that timeslices are attached to one another only by edges between consecutive copies of the same node. The interpretation is that this represents a network in which the links are memoryless; only nodes can retain information from one timeslice to the next.

Definition 30. A *data payload sheaf* $\mathcal{D}$ over a time-dependent link complex L with nodes N is parameterized by

(1) A vector space D of possible packets, and
(2) A transmit queue length $n - 1$.

The stalks of $\mathcal{D}$ are given by

For each vertex c of L: $(\{a \in N : \text{there exists a cell } d \in X \text{ with } c \subset d \text{ and } a \in d\} \cup \{\perp\})^2 \times D^n$

For each edge of the form $((c, t), (c, t+1))$: $(\{a \in N : \text{there exists a cell } d \in X \text{ with } c \subset d \text{ and } a \in d\} \cup \{\perp\}) \times D^{n-1}$

For all other simplices c of L: $(\{a \in N : \text{there exists a cell } d \in X \text{ with } c \subset d \text{ and } a \in d\} \cup \{\perp\}) \times D$

The restrictions of $\mathcal{D}$ are given by

(1) Between timeslices

$$\mathcal{D}\left((a,t) \subset ((a,t),(a,t+1))\right)(n_1,n_2,x_1,\dots,x_n) = \begin{cases} (n_2,x_2,\dots,x_{n-1},0) & \text{if } n_2 = a \text{ and } x_n \neq 0 \\ (n_2,x_{p(1)},\dots,x_{p(n-1)}) & \text{if } n_2 \neq a \text{ and } n_2 \neq \perp \\ (\perp,x_2,\dots,x_n) & \text{otherwise} \end{cases}$$

where $(x_1,\dots,x_n) \mapsto (x_{p(1)},\dots,x_{p(n-1)})$ is a receive queue function.

$$\mathcal{D}\left((a,t+1) \subset ((a,t),(a,t+1))\right)(n_1,n_2,x_1,\dots,x_n) = \begin{cases} (n_1,x_3,\dots,x_n,0) & \text{if } n_1 = a \\ (n_1,x_2,\dots,x_n) & \text{if } n_1 \neq a \text{ and } n_1 \neq \perp \\ (\perp,x_2,\dots,x_n) & \text{otherwise} \end{cases}$$

(2) Within timeslices, all restrictions between simplices $a \subset b$ of dimension 1 or higher are of the form

$$\mathcal{D}\left((a,t) \subset (b,t)\right)(n,x) = \begin{cases} (n,x) & \text{if } (n,x) \in \mathcal{D}((b,t)) \\ (\perp,0) & \text{otherwise} \end{cases}$$

while restrictions from a vertex a to an edge (a,b) are given by

$$\mathcal{D}\left((a,t) \subset ((a,t),(b,t))\right)(n_1,n_2,x_1,\dots,x_n) = \begin{cases} (n_2,x_n) & \text{if } n_2 = a \text{ and } x_n \neq 0 \\ (n_2,x_1) & \text{if } n_2 \neq a \\ (\perp,0) & \text{otherwise} \end{cases}$$

Proposition 31. *Every data payload sheaf contains an activation sheaf as a subsheaf when restricted to any timeslice.*

This means that the data payload sheaf incorporates the transmission structure described previously for activation sheaves, and more importantly that the within-timeslice restrictions for the data payload sheaf describe the relationship between the data on links and within the nodes.

For clarity, if $\mathcal{D}$ is a data payload sheaf on a time-dependent link complex X, the activation subsheaf at time t is written $A_t\mathcal{D}$. Therefore, there is a collection of surjections on stalks $A_t(a) : \mathcal{D}(a,t) \to A_t\mathcal{D}(a)$ that project out the appropriate components of the stalks. These surjections have the property that they are compatible with both sheaves, in that if $a \subset b$

$$A_t(b) \circ \mathcal{D}(a \subset b) = A_t\mathcal{D}(a \subset b) \circ A_t(a).$$

This is taken to be the description of a *sheaf morphism* $A_t : \mathcal{D} \to A_t\mathcal{D}$.

Proposition 32. *When restricted to a single node n, every data payload sheaf contains a 2-grouping sheaf taking values in the nodes adjacent to n as a subsheaf.*

As a result, transmissions between timeslices are decoupled from one another. It is important to realize that this does *not* mean that the data payloads are decoupled. Instead, given a sequence of nodes that transmit at each timeslice – global sections of each activation sheaf in each timeslice – the data payload sheaf will describe the pathways for threading data through the network as the next proposition states.

Proposition 33. *Given a time-dependent link complex X and a data payload sheaf $\mathcal{D}$ and global sections $\{s_t\}$ for each activation subsheaf $A_t\mathcal{D}$, then*

(1) *the restriction of each stalk $\mathcal{D}(a,t)$ to the collection of elements whose image through $A_t(a)$ is $s_t(a)$ yields a subsheaf $\mathcal{P}$ and*

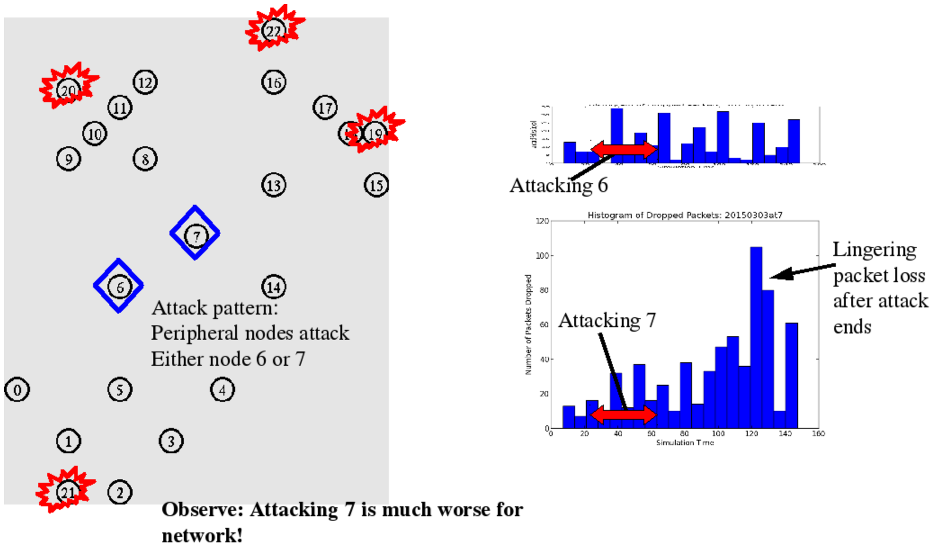


FIGURE 12. Packet drop results for the two attacks in the 20150303 network; notice the substantial difference depending on the target

(2) *$\mathcal{P}$ is a sheaf taking values in the same category as the data payloads D, and*
(3) *if D is a vector space, the dimension of the space of global sections of $\mathcal{P}$ is an upper bound on the network's throughput given the transmission pattern described by $\{s_t\}$.*

Beware that the upper bound given in Proposition 33 is not tight – there may be global sections that describe packets that do not reach their intended destination(s).

4.3. **Local homology detects vulnerable nodes.** It is known [35] that the target node in an attack has a significant impact on the response of the network. As a simple example, consider Figure 12, which shows the packets dropped as a function of time for two attacks on a network that were identical except for their target. As the Figure makes clear, an attack on Node 7 is much more determinental to the network. Figure 13 shows the local homology dimension LH_1 over the faces of the link complex of the network. Notice that Node 7 is in the interior of a region with higher LH_1 than Node 6, which we take to be indicative of a local vulnerability. In particular, according to Theorem 18, an attack on Node 7 splits the network into more connected components than an attack on Node 6. This means that more nodes are unable to communicate while Node 7 is overloaded. In Section 4.5, we explain that because LH_1 correlates with the number of forwarded packets, it is a good measure of node (and even link) vulnerability.

4.4. **Forwarded packet distribution.** In the process of searching for a method for quantifying the topological features of network traffic, we asked the question "How does moving the traffic around in the network (especially near the topological boundary) affect forwarding patterns?" Although we rely on local homology for the determination of boundaries in a topological space (Figure 3), we wanted

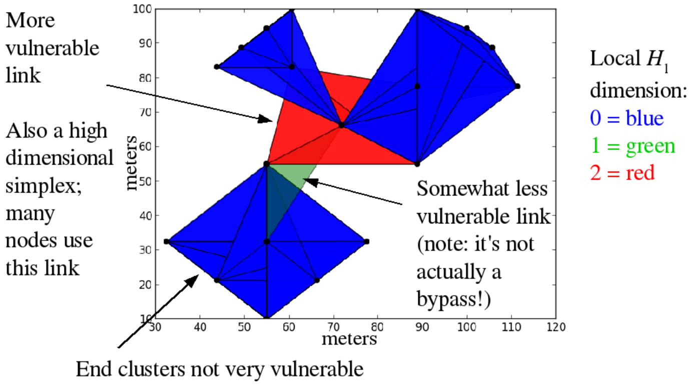

More vulnerable link
Also a high dimensional simplex; many nodes use this link
Somewhat less vulnerable link (note: it's not actually a bypass!)
End clusters not very vulnerable
Local H_1 dimension:
0 = blue
1 = green
2 = red
meters
meters
100
90
80
70
60
50
40
30
20
10
30
40
50
60
70
80
90
100
110
120

FIGURE 13. Local homology dimension for the 20150303 network, explaining why node 7 is more vulnerable

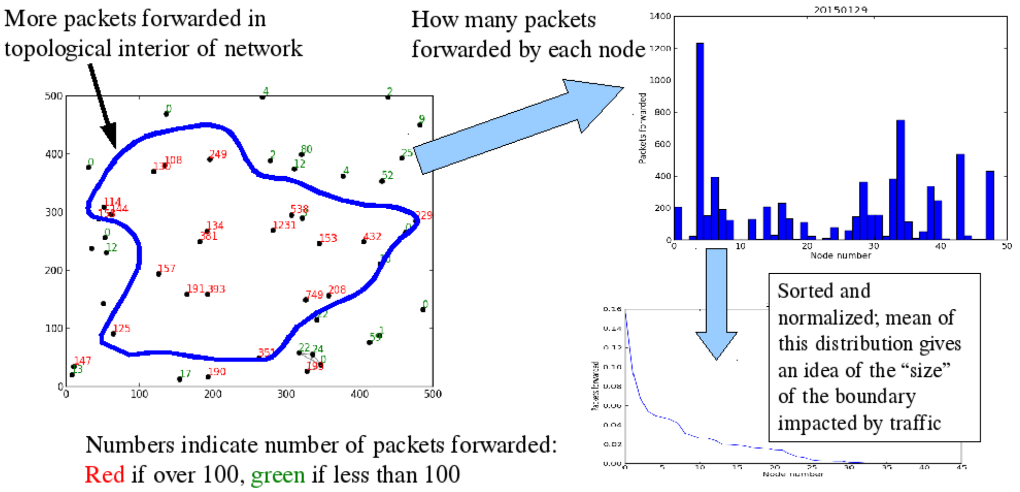


FIGURE 14. Process for computing forwarded packet distributions.

an independent measurement based on packet statistics. Therefore, we wondered if there was a connection between the number of packets forwarded by a given node and its position in the network. Along the way, we wondered if the distribution of these forwarded packet counts would be useful in its own right. Therefore, we performed the analyses diagrammed in Figure 14 to produce a distribution of forwarded packets as a function of nodes. To control for irrelevant differences caused by node labeling, we sorted the packet counts from greatest to least. One can interpret the forwarded packet count in terms of the conditional probability that given a packet will be forwarded, what is the probability that a particular node will be given the task of forwarding it. Clearly, this distribution is dependent on the exact traffic patterns being carried by the network, because the desired source

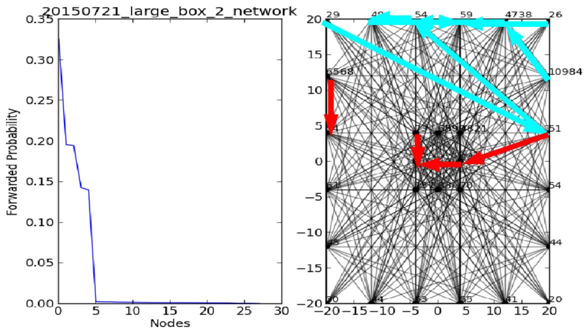


FIGURE 15. Forwarded packet distribution is skewed left for low traffic

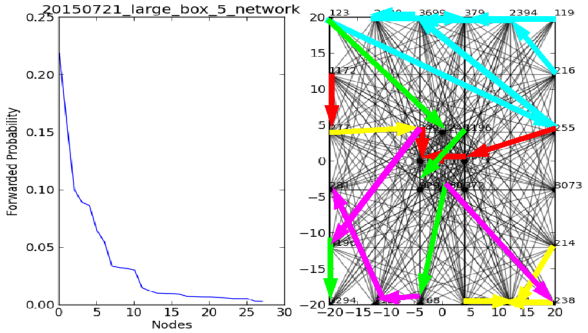


FIGURE 16. Forwarded packet distribution with progressively more traffic

and destinations will clearly impact which nodes forward any given packet. This variability can be seen in Figures 15 – 17. As the traffic through the interior of the network increases, the distribution broadens because more nodes are required (especially as the queues in some of the more opportune nodes fill). Due to the presence of several competing effects (at least network topology, specific traffic conditions, queue sizes), we suggest further study of the forwarded packet distribution is required to better understand how to interpret it.

4.5. **Local homology correlates with forwarded packets.** The structure of the global sections of an activation sheaf leads to a model in which an active node silences all other nodes in its vicinity.

Definition 34. Because of Lemma 25 and 26, we call the star over an active region associated to a node n the *region of influence*. The region of influence of a facet is the star over the closure of that facet. The region of influence for a collection of facets F can be written as a union

$$\text{roi } F = \bigcup_{f \in F} \text{star cl } f.$$

The following Corollary indicates the correct region to use in computing local homology in a wireless network.

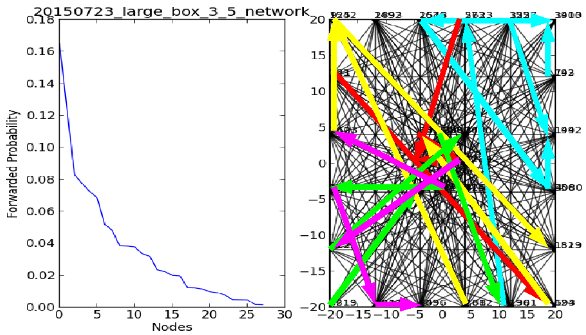


FIGURE 17. Forwarded packet distribution with substantially more traffic

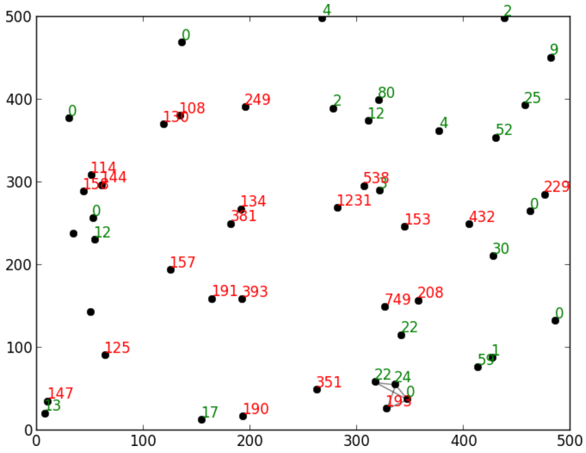


FIGURE 18. Locations of nodes and forwarded packet counts (axes in meters) in 20150129

Corollary 35. *The complement of the region of influence of a facet is a closed subcomplex.*

Given this justification, [33] shows that critical nodes or links are those cells c for whom the local homology dimension (see also [24])

$$LH_k(c) = \dim H_k(X, X \setminus \text{roi } c)$$

is larger than the average.

This implies the following experimental hypothesis: *If a node is critical, it will have a large local homology dimension.* Since the *ns*2 network simulator provides complete transcripts of all packets, we can define a critical node to be one that *forwards* a large number of packets compared to other nodes in the network [3].

We constructed a small simulation with 50 nodes as shown in Figure 18. Packets were randomly assigned source and destination nodes within the network, and all packet histories were recorded for analysis.

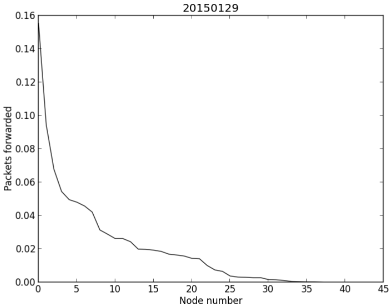


FIGURE 19. Probability that a given packet will be forwarded by a specific node for the 20150129 dataset (compare Figures 15 – 17)

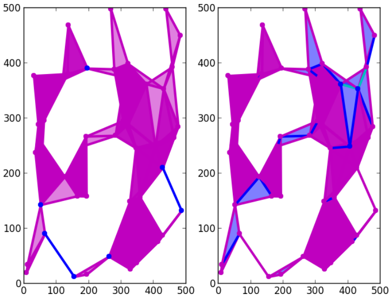


FIGURE 20. Dimension of local homology LH_1 (left) and LH_2 (right) for the 20150129 network. Axes in meters; Magenta = 0, Blue = 1, Cyan = 2.

Figure 19 shows the probability that a node will forward a random packet. (The node numbers have been sorted from greatest to least probability.) The figure shows that most nodes forward only a small number of packets, while a few nodes carry considerably more traffic.

Figure 20 shows the dimension of local homology over all nodes and links in the network. In this particular network, the local homology dimension is only 0, 1, or 2. It is clear that nodes with high LH_1 occupy certain "pinch points" in the network.

Figure 21 shows the probability that a node forwarding a certain number of packets will have the given value of LH_1. (We did not find a strong correspondence between forwarded packets and LH_2.) It is immediately clear that all nodes forwarding a large number of packets are assigned a high local homology, but the converse is not necessarily true. Local homology dimension is an indication that a node may be critical, but does not guaranteed that it actually is.

4.6. **Relative simplicial homology library.** As part of this project and several others our team is pursuing, we found it appropriate to develop a general simplicial homology library that supports relative and local homology computations. It appears that this is the first ever library[3] that supports relative homology. Since

[3] We even asked the authors of [5] (the only other applied local homology paper we know of), and they know of no such library available.

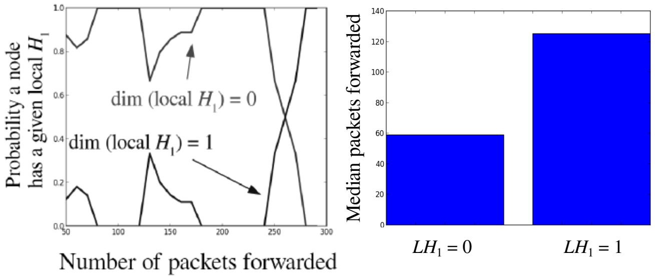


Figure 21. Probability a node has a certain local homology dimension given the number of packets it forwards for the 20150129 dataset

this library has wide applicability across programs – and is purely mathematics – we are pursuing an open source development model. The most recent version is available at

`https://github.com/kb1dds/pysheaf`

This library is written in Python 2.7.3 and uses the `NetworkX` and `numpy` python libraries. This library also includes a battery of tests (including those shown in Figure 3 and 4). The library operates on abstract simplicial complexes (Section 3.1). It supports both full representations (in which all simplices are stored) and toplex representations (which includes only the maximal simplices).

4.7. **New conjecture about wireless network tomography.** Near the end of the program, we realized that it would become important to study the connection between network topology and geometry. One of our colleagues at American University (Prof. Stephen Casey) pointed out that network tomography appears to be a form of a combinatorial Radon transform, when a network is endowed with a geometry. He observed that the Radon transform, which takes the interior values of a function on a Riemannian manifold to its boundary, is injective for hyperbolic spaces. If a network had the analogous hyperbolic geometry, then it might make the combinatorial Radon transform injective as well. If the function on the network were the counts of packets forwarded, then an injective transform would mean that this distribution could be reconstructed from values at the periphery of the network. This would mean that under certain circumstances, packet statistics over the entire network could be recovered from a few choice locations.

With the help of Danielle Beard (a master's student working jointly with Prof. Robinson and Prof. Casey), we formulated the following conjecture:

Conjecture 36. *The combinatorial Radon transform is injective when the local curvature of the network is negative over the entire network.*

We conjecture that a cellular version of the Gauss-Bonnet theorem applies to networks, and would therefore provide conditions under which lossless network tomography is possible. A convenient starting place is our earlier work [9] on a microlocal Gauss-Bonnet theorem. It will be necessary to specify geometric information on the network, which could be induced from spatial locations, from the hop-length metric, or something more general. The induced metric from a spatial embedding is likely to be the easiest to test, as `ns2` can simulate such a network directly. We can then test to see (1) if network paths tend to follow geodesics and (2) if they tend to converge, as would be expected in a space with negative curvature.

A future direction for study would be to generate simulations of networks in which the local curvature and the local homology dimension varies. Unfortunately, since `ns2` simulates networks purely in Euclidean space (and therefore *not* in hyperbolic space), it quite difficult to generate networks that are both embedded in Euclidean space and have negative curvature. Ideally, we would like to vary the curvature bound to see if geometric effects become more prominent. We attempted to construct, but did not complete, a family of embedded networks with a variable negative curvature bound.

5. Important findings and conclusions

This project was centered around the hypothesis that *higher-dimensional network topology* plays an important role in determining vulnerabilities of an *ad hoc* wireless network. Contrary to the focus of much of the network vulnerability literature, while network protocol is important, it is not the only determinant of performance. Since we employed a simulator `ns2` that does not model signal degradation, we were able to isolate purely topological phenomena in wireless network behavior. We showed that *persistent homology* is a coarse, but effective assessment of the overall vulnerability of a network to specific kinds of network attacks, but did not pursue a systematic study of network attack and defence strategies.

On this project, we discovered that the *distribution of forwarded packets* over the nodes of a network is a good way to identify network vulnerabilities. Unfortunately, measuring the forwarded packet distribution requires access to packet statistics at every node in the network, which limits its direct practical application. We found that *local homology dimension* in degree 1 correlates with forwarded packet counts, and can be measured from the topology of the network. We expect that local homology can be computed more easily from a network, since it only requires knowledge of the 2-hop neighborhood of a node. This neighborhood is justified by the use of *activation sheaves* to model CSMA/CD protocols, which are typically used in *ad hoc* wireless networks. We showed that activation sheaves describe useful local neighborhoods of nodes, but little else. Finally, we showed that activation sheaves are merely the first in a heirarchy of *data payload sheaves* that model the history of packets in a network, and therefore likely encapsulate measures of its capacity.

5.1. Publication list.

The following publications resulted directly from this program:

(1) Technical report: "Protocol-independent critical node detection" (Submitted to prepublication review 27 May 2016)
(2) Technical report: "Modeling wireless network routing using sheaves" (Submitted to prepublication review 28 May 2016)
(3) Talk: "Sheaf-based modeling of wireless communications" (DISTAR 24257)
(4) Talk: "Sheaf-based communication network invariants" (DISTAR 25664)

The following publications in support of our team's other programs are relevant to the objectives of this program:

(1) "Sheaf and cosheaf methods for analyzing multi-model systems," `arXiv:1604.04647`.
(2) Danielle Beard, "Network tomography and the Radon transform," Master's thesis, May 2016.
(3) Cliff Joslyn, Brenda Praggastis, Emilie Purvine, Arun Sathanur, Michael Robinson, Stephen Ranshous, "Local Homology Dimension as a Network Science Measure," accepted to *SIAM Workshop on Network Science* 2016, July 15-16, 2016, Boston.
(4) Emilie Purvine, Michael Robinson, and Cliff Joslyn, "Categorification in the real world." *Joint Mathematics Meetings MAA Session on Mathematics Experiences and Projects in Business, Industry, and Government*, Seattle, WA. January 8, 2016.
(5) "DARPA Tutorial on Sheaves in Data Analytics", American University, Washington, DC, August 25-26, 2015. `http://www.drmichaelrobinson.net/sheaftutorial/index.html`

6. Significant hardware development

Not applicable.

7. Special comments

Not applicable.

8. Implications for future research

This project has opened three main avenues for further topological studies of *ad hoc* wireless networks.

8.1. **Network simulations.**

The first class of questions involve additional and more detailed network simulations.

(1) Now that we have established the importance of topological effects on *ad hoc* wireless networks, it is important to assess the relative importance between network topology, network geometry, and protocol effects. Studies should be developed to hold the link complex fixed while varying geometry or protocol parameters, to isolate those effects from topological effects. Because of the potential importance of an injective combinatorial Radon transform for network topology, the first such study should be to analyze the impact of network curvature on packet statistics.
(2) Since the distribution of forwarded packets appears to be both novel and useful, we advocate performing larger statistical studies to tease apart the competing effects.
(3) How does the local and global topology impact a network's response to a burst of traffic? We expect that the presence of nontrivial loops in the link complex will allow a network to return to its quiescent state more quickly, as there are alternate paths for packets to be dispersed.
(4) Related to this question is how changes in network topology while the network is running impact its performance. This could allow our techniques to be used to assess the deployment of different adaptive capacity management strategies.

(5) Finally, although we tested a few specific attack patterns, we advocate for a systematic topological study of adversarial network strategies.

8.2. **Prove conjectures.** We have posed a number of mathematical conjectures about the models we developed on this project. Future studies could attempt to prove these conjectures and study their implications:

(1) Is the combinatorial Radon transform injective when the curvature is negative? If so, this would enable network tomography based on the packet statistics of a small part of the network.
(2) How are packet histories embedded in a data payload sheaf? More importantly, how does one extract the capacity of a wireless network from its description as a data payload sheaf?
(3) What is lost in using only netflow statistics? These are aggregate counts of packets being sent between nodes, without an indication of their routing or whether they have forwarded or not. Although the netflow statistics are routinely gathered, our analysis indicates that knowing the routing pattern is important for assessing network performance and vulnerabilities.

8.3. **Numerical experiments with sheaves.** Finally, we propose that additional mathematical aspects of sheaves could be explored numerically, to form additional useful conjectures.

(1) We performed our analyses offline because of the computational load. If topological analysis is to be deployed in practice, it ought to be performed for larger networks and ought to be computed much quicker. Therefore, it is necessary to improve the computational scaling of topological computations. Although some aspects of our analyses can be parallelized (computing the local homology dimension requires knowledge of a small neighborhood and is independent of other neighborhoods), others (the homology computation itself) have been stubbornly resistant to parallelization.
(2) Studies could explore cohomology of the various sheaf encodings we developed. We know activation sheaves don't have interesting cohomology, but what about data payload sheaves?

References

[1] The NS-2 network simulator. http://www.nsnam.org/. Accessed: 2016-05-23.

[2] N. Ahmed, S. S. Kanhere, and S. Jha. The holes problem in wireless sensor networks: a survey. *SIGMOBILE Mobile Computing and Communications Review*, 9(2):4–18, 2005.

[3] Ashwin Arulselvan, Clayton W Commander, Lily Elefteriadou, and Panos M Pardalos. Detecting critical nodes in sparse graphs. *Computers & Operations Research*, 36(7):2193–2200, 2009.

[4] Amir Bashan, Yehiel Berezin, Sergey V. Buldyrev, and Shlomo Havlin. The extreme vulnerability of interdependent spatially embedded networks. *Nature Physics*, 9:667–672, 2013.

[5] P. Bendich, D. Cohen-Steiner, H. Edelsbrunner, J. Harer, and D. Morozov. Inferring local homology from sampled stratified spaces. In *Foundations of Computer Science (FOCS)*, pages 536–546, Providence, RI, 2007.

[6] M. Chiang, S. Low, R. Calderbank, and J. Doyle. Layering as optimization decomposition:a mathematical theory of network architectures. *Proc. IEEE*, 95(1), January 2007.

[7] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. *Discrete and Computational Geometry*, 37(1):103–120, 2007.

[8] C. W. Commander, P. Pardalos, V. Ryabchenko, S. Uryasev, and G. Zrazhevsky. The wireless network jamming problem. *J. Comb. Optim.*, 14:481–498, 2007.

[9] J. Curry, R. Ghrist, and M. Robinson. Euler calculus and its applications to signals and sensing. In Afra Zomorodian, editor, *Proceedings of Symposia in Applied Mathematics: Advances in Applied and Computational Topology*. 2012.

[10] Vin de Silva and Robert Ghrist. Coverage in sensor networks via persistent homology. *Algebraic & Geometric Topology*, 7(339-358):24, 2007.

[11] Vin De Silva, Robert Ghrist, and Abubakr Muhammad. Blind swarms for coverage in 2-d. In *Robotics: Science and Systems*, pages 335–342, 2005.

[12] Marco Di Summa, Andrea Grosso, and Marco Locatelli. Complexity of the critical node problem over trees. *Computers & Operations Research*, 38(12):1766–1774, 2011.

[13] Thang N Dinh, Ying Xuan, My T Thai, Panos M Pardalos, and Taieb Znati. On new approaches of assessing network vulnerability: hardness and approximation. *Networking, IEEE/ACM Transactions on*, 20(2):609–619, 2012.

[14] Gaogao Dong, Jianxi Gao, Lixin Tian, Ruijin Du, and Yinghuan He. Percolation of partially interdependent networks under targeted attack. *Physical Review E*, 85, 2012.

[15] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discrete and Computational Geometry*, 28:511–533, 2002.

[16] Bernard Fortz, Jennifer Rexford, and Mikkel Thorup. Traffic engineering with traditional ip routing protocols. *Communications Magazine, IEEE*, 40(10):118–124, 2002.

[17] Jianxi Gao, Sergey Buldyrev, Shlomo Havlin, and H. Eugene Stanley. Robustness of a network of networks. *Physics Review Letters*, 107, November 2011.

[18] R. Ghrist. Barcodes: the persistent topology of data. *Bulletin-American Mathematical Society*, 45(1):61, 2008.

[19] R. Ghrist and Y. Hiraoka. Applications of sheaf cohomology and exact sequences to network coding. *preprint*, 2011.

[20] R. Ghrist and A. Muhammad. Coverage and hole-detection in sensor networks via homology. In *Proceedings of the 4th international symposium on Information processing in sensor networks*, page 34. IEEE Press, 2005.

[21] Assane Gueye, Jean C Walrand, and Venkat Anantharam. Design of network topology in an adversarial environment. In *Decision and Game Theory for Security*, pages 1–20. Springer, 2010.

[22] A. Hatcher. *Algebraic Topology*. Cambridge University Press, 2002.

[23] K. Jain, J. Padhye, V. Padmanabhan, and L. Qiu. Impact of interference on multi-hop wireless network performance. In *Proc. ACM MobiCom*, 2003.

[24] Cliff Joslyn, Brenda Praggastis, Emilie Purvine, Arun Sathanurand Michael Robinson, and Stephen Ranshous. Local homology dimension as a network science measure. In *accepted to SIAM Workshop on Network Science 2016*, Boston, July 2016.

[25] Jinko Kanno, Jack G Buchart, Rastko R Selmic, and Vir Phoha. Detecting coverage holes in wireless sensor networks. In *Control and Automation, 2009. MED'09. 17th Mediterranean Conference on*, pages 452–457. IEEE, 2009.

[26] J.-W. Lee, M. Chiang, and R. Calderbank. Utility-optimal random-access control. *IEEE Trans. Wireless Comm.*, 6(7):2741–2751, 2007.

[27] John Milnor and James D. Stasheff. *Characteristic Classes*. Princeton University Press, 1974.

[28] K. Mischaikow and V. Nanda. Morse theory for filtrations and efficient computation of persistent homology. *Discrete and Computational Geometry*, 50(2):330–353, 2013.

[29] J. Munkres. *Elements of Algebraic Topology*. Westview Press, 1984.

[30] T. Nandagopal, T.-E. Kim, X. Gao, and V. Bharghavan. Achieving mac layer fairness in wireless packet networks. In *Proc. ACM MobiCom*, pages 87–98, 2002.

[31] Guevara Noubir. On connectivity in ad hoc networks under jamming using directional antennas and mobility. In *Wired/Wireless Internet Communications*, pages 186–200. Springer, 2004.

[32] M. Robinson. Asynchronous logic circuits and sheaf obstructions. *Electronic Notes in Theoretical Computer Science*, pages 159–177, 2012.

[33] Michael Robinson. Analyzing wireless communication network vulnerability with homological invariants. In *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Atlanta, Georgia, 2014.

[34] Toshihiro Tanizawa, Shlomo Havlin, and H. Eugene Stanley. Robustness of onionlike correlated networks against targeted attacks. *Physical Review E*, 85, 2012.

[35] Walter Willinger, David Alderson, and John C. Doyle. Mathematics and the internet: A source of enormous confusion and great potential. *Notices of the AMS*, 56(5), May 2009.

[36] W. Xu, W. Trappe, Y. Zhang, and T. Wood. The feasibility of launching and detecting jamming attacks in wireless networks. In *Proc. ACM MobiHoc*, 2005.

[37] Wenyuan Xu, Ke Ma, Wade Trappe, and Yanyong Zhang. Jamming sensor networks: attack and defense strategies. *Network, IEEE*, 20(3):41–47, 2006.

[38] X. Yang and N. Vaidya. Priority scheduling in wireless ad hoc networks. In *Proc. ACM MobiHoc*, 2002.

[39] Jixing Yao, Guyu Zhang, Jinko Kanno, and Rastko Selmic. Decentralized detection and patching of coverage holes in wireless sensor networks. In *SPIE Defense, Security, and Sensing*, pages 73520V–73520V. International Society for Optics and Photonics, 2009.

Department of Mathematics and Statistics, American University, 4400 Massachusetts Ave NW, Washington, DC 20016

E-mail address: `michaelr@american.edu`